

Servicios de Nube - Azure 01

Bases de Datos Avanzadas

Universidad del Rosario

Maestría MACC

1 Azure

- Propiedades de Azure

2 Azure Storage Explorer

3 Data Lake

- Tratamiento a Grandes Volúmenes de Datos
- Diferencias con Data Warehouse

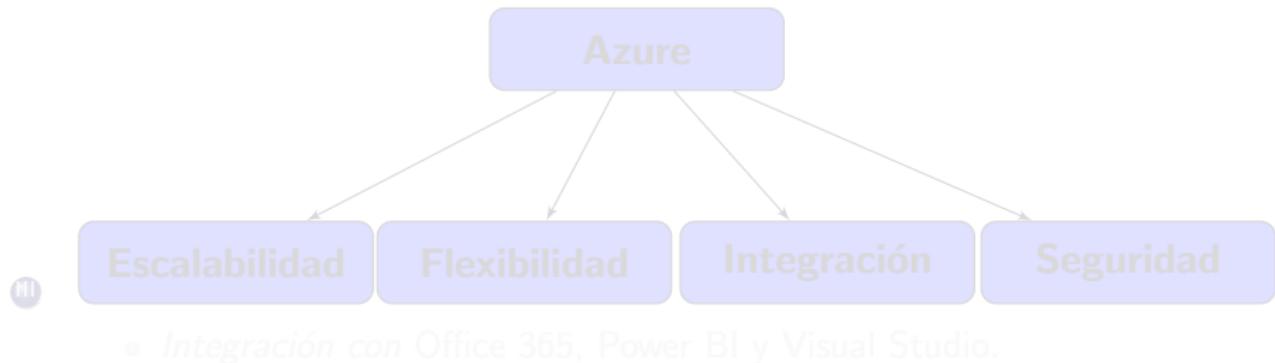
4 Azure Data Factory

- Usos de Data Factory

Características



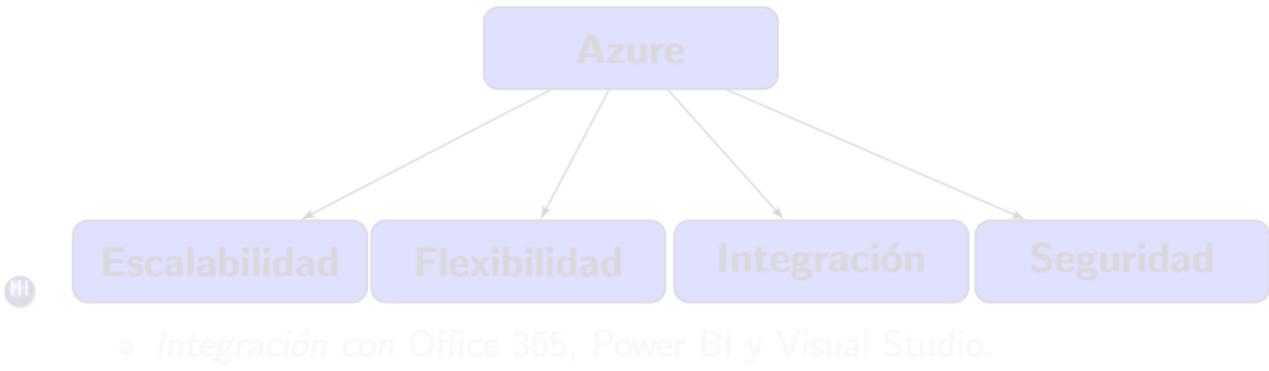
- ① Servicio tipo nube/*cloud*.
- ① Creación e implementación de aplicaciones y servicios.



Características



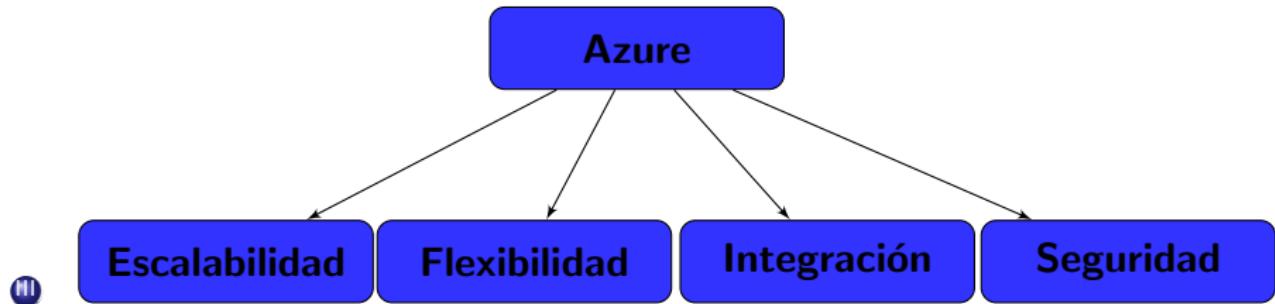
- ➊ Servicio tipo nube/*cloud*.
- ➋ Creación e implementación de aplicaciones y servicios.



Características

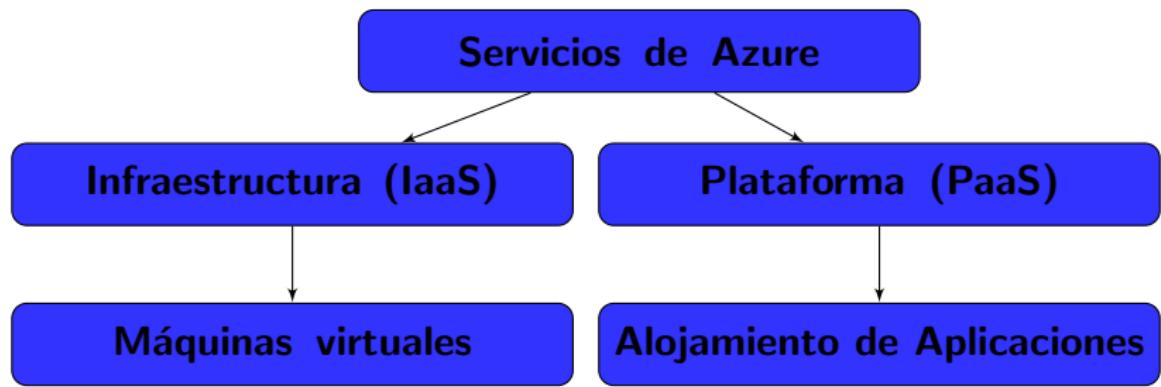


- ➊ Servicio tipo nube/*cloud*.
- ➋ Creación e implementación de aplicaciones y servicios.



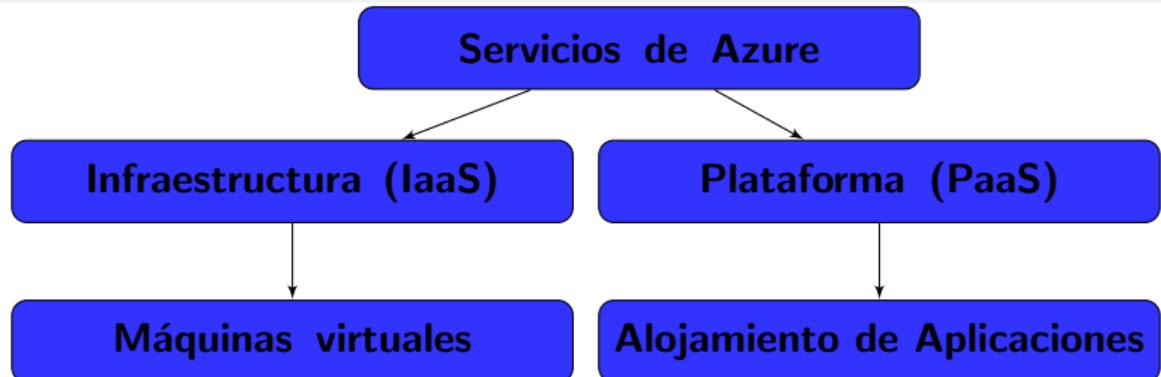
- *Integración con Office 365, Power BI y Visual Studio.*

Escalabilidad

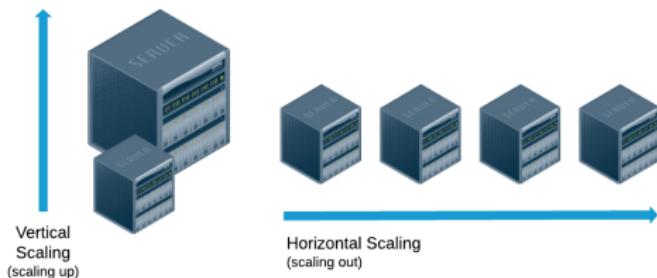


- Tipos de escalabilidad: *horizontal y vertical.*

Escalabilidad

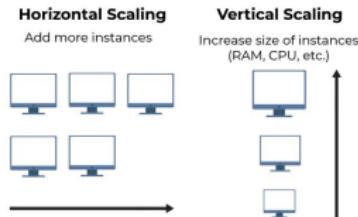


- II Tipos de escalabilidad: *horizontal* y *vertical*.



Servicios para Escalabilidad Horizontal y Vertical

Horizontal Scaling vs. Vertical Scaling



① Horizontal:

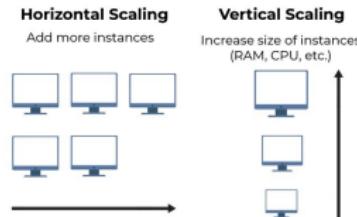
- Azure Virtual Machines, Azure Kubernetes Service (AKS), Service Fabric, Azure App Service.
- *Agregar más servidores.*

② Vertical:

- Azure VM Scale, Azure SQL Database.
- *Potenciar un servidor.*

Servicios para Escalabilidad Horizontal y Vertical

Horizontal Scaling vs. Vertical Scaling



① Horizontal:

- Azure Virtual Machines, Azure Kubernetes Service (AKS), Service Fabric, Azure App Service.
- *Agregar más servidores.*

② Vertical:

- Azure VM Scale, Azure SQL Database.
- *Potenciar un servidor.*

Flexibilidad y Seguridad

① Flexibilidad:

② Servicios de Plataforma como servicio (PaaS).

- .NET, Java, Python, Node.js, MySQL, etc.

③ Servicios de Infraestructura (IaaS).

- VMs, contenedores y servicios de aplicaciones.

④ Seguridad:

- Control de acceso basado en roles (RBAC).
- Encriptación, seguridad de red.

⑤ Herramientas adicionales:

- Azure Key Vault, Azure Active Directory, etc.

Flexibilidad y Seguridad

I Flexibilidad:

II Servicios de Plataforma como servicio (PaaS).

- .NET, Java, Python, Node.js, MySQL, etc.

III Servicios de Infraestructura (IaaS).

- VMs, contenedores y servicios de aplicaciones.

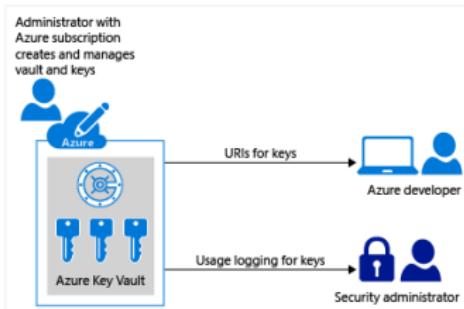
IV Seguridad:

- Control de acceso basado en roles (RBAC).
- Encriptación, seguridad de red.

V Herramientas adicionales:

- **Azure Key Vault**, Azure Active Directory, etc.

Seguridad: Azure Key Vault



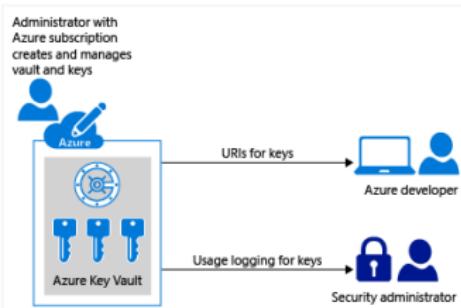
I Administración y creación de claves y secretos.

- *Secreto:* información confidencial usada en una aplicación o servicio (p. ej., certificado de cliente).

II Características:



Seguridad: Azure Key Vault



I Administración y creación de claves y secretos.

- *Secreto:* información confidencial usada en una aplicación o servicio (p. ej., certificado de cliente).

II Características:



Propiedades del Programa



- ① Interacción con almacenamiento en nube (p. ej., Blob Storage).
- ② Características:



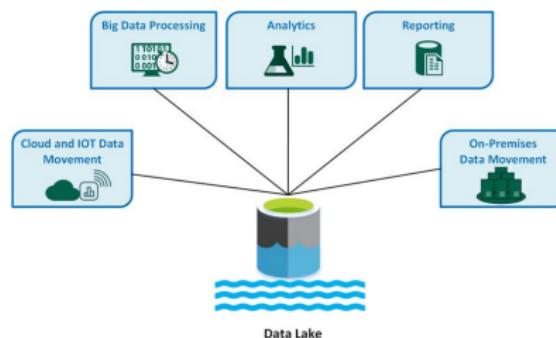
Propiedades del Programa



- ① Interacción con almacenamiento en nube (p. ej., Blob Storage).
- ② Características:

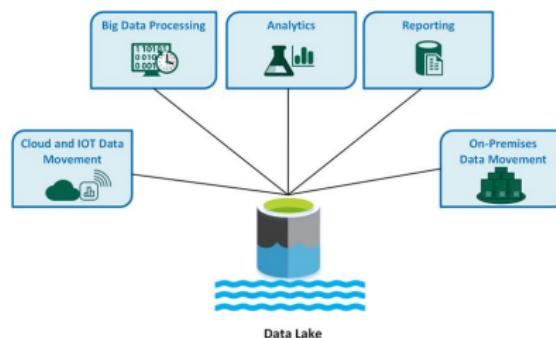


Azure Data Lake



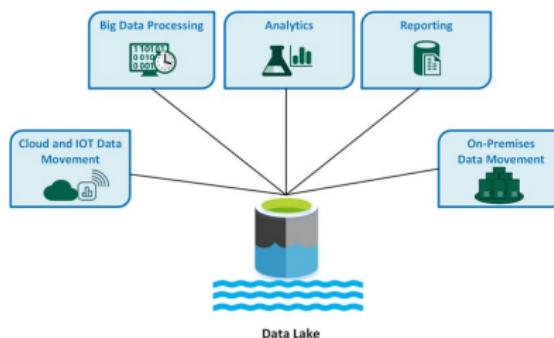
- ① Almacenamiento (Data Lake Storage) y análisis de datos (Data Lake Analytics).
 - *Estructurados y No Estructurados*
- ② Tratamiento de grandes volúmenes en paralelo.
- ③ Lenguajes compatibles:
 - U-SQL (SQL + C#), Python, R y Java.

Azure Data Lake



- ① Almacenamiento (Data Lake Storage) y análisis de datos (Data Lake Analytics).
 - *Estructurados y No Estructurados*
- ② Tratamiento de grandes volúmenes en paralelo.
- ③ Lenguajes compatibles:
 - U-SQL (SQL + C#), Python, R y Java.

Azure Data Lake



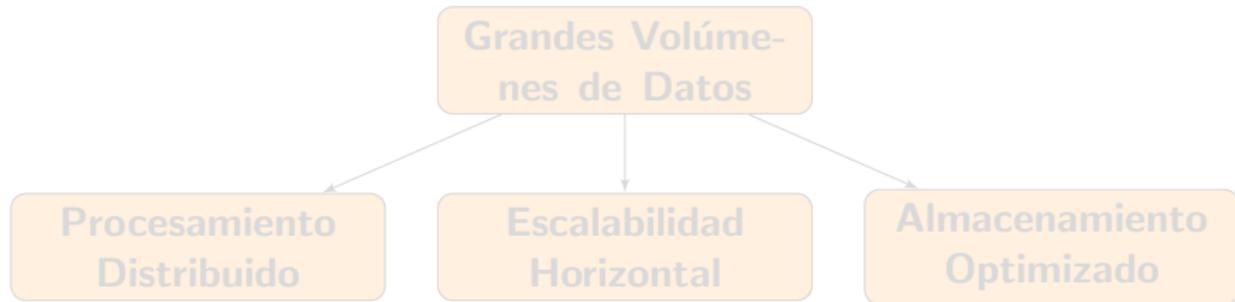
- ① Almacenamiento (Data Lake Storage) y análisis de datos (Data Lake Analytics).
 - *Estructurados y No Estructurados*
- ② Tratamiento de grandes volúmenes en paralelo.
- ③ Lenguajes compatibles:
 - U-SQL (SQL + C#), Python, R y Java.

Usos y Grandes Volúmenes de Datos

① Usos de Data Lake:

- Almacenamiento y análisis de grandes volúmenes de datos.
- Sistemas de análisis en tiempo real.
- Creación de aplicaciones para ML.

11

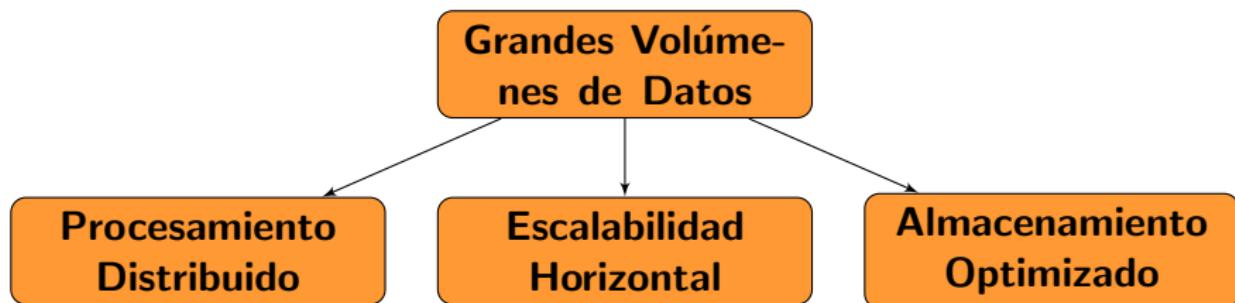


Usos y Grandes Volúmenes de Datos

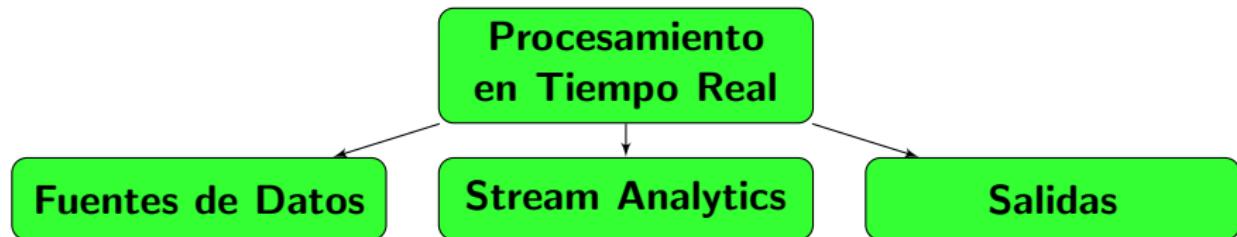
① Usos de Data Lake:

- Almacenamiento y análisis de grandes volúmenes de datos.
- Sistemas de análisis en tiempo real.
- Creación de aplicaciones para ML.

②



Procesamiento en Tiempo Real



- Fuentes: Blob Storage, Data Lake Storage, Event Hubs, etc.
- Salidas: Base SQL, Blob Storage, etc.

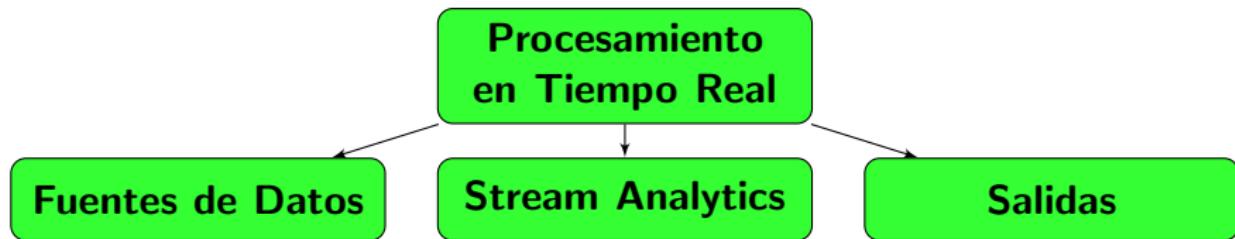


Blob Storage:

- Almacenamiento de datos tipo documentos, imágenes, video...
- Altamente escalable.
- Integración con otros servicios de Azure.

Procesamiento en Tiempo Real

11



- Fuentes: Blob Storage, Data Lake Storage, Event Hubs, etc.
- Salidas: Base SQL, Blob Storage, etc.



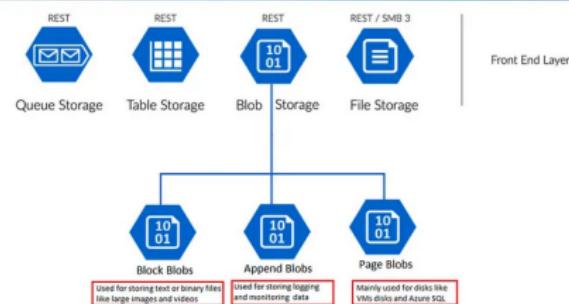
11

Blob Storage:

- Almacenamiento de datos tipo documentos, imágenes, video...
- Altamente escalable.
- Integración con otros servicios de Azure.

Procesamiento en Tiempo Real: Blob Storage

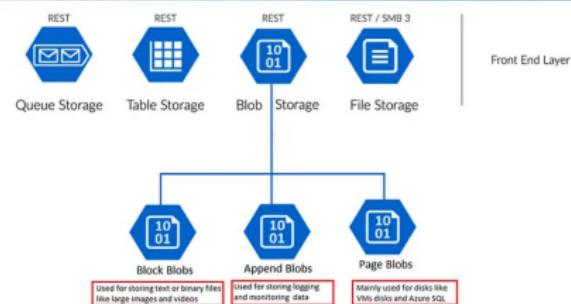
Azure Storage Architecture



- ① Acceso mediante URL única.
- ② Altamente escalable.
- ③ Integrable con Azure Data Factory.
- ④ Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento en Tiempo Real: Blob Storage

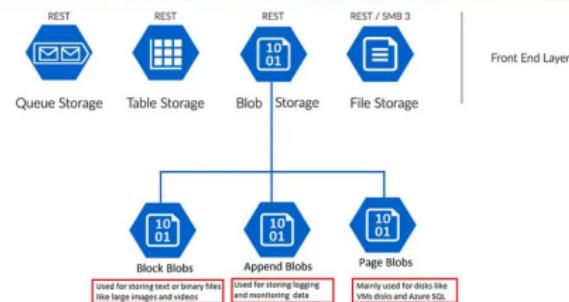
Azure Storage Architecture



- ① Acceso mediante URL única.
- ② Altamente escalable.
- ③ Integrable con Azure Data Factory.
- ④ Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento en Tiempo Real: Blob Storage

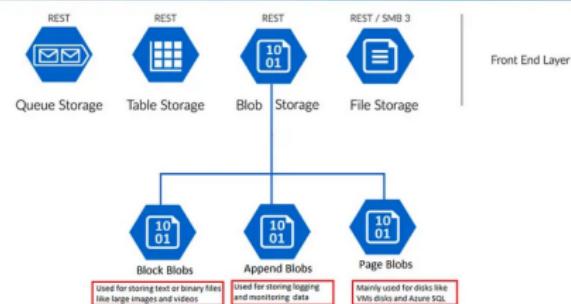
Azure Storage Architecture



- ➊ Acceso mediante URL única.
- ➋ Altamente escalable.
- ➌ Integrable con Azure Data Factory.
- ➍ Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento en Tiempo Real: Blob Storage

Azure Storage Architecture



- I Acceso mediante URL única.
- II Altamente escalable.
- III Integrable con Azure Data Factory.
- IV Características avanzadas:
 - Ciclo de vida en datos.
 - Cifrar datos en reposo y tránsito.
 - Control de acceso.

Procesamiento de Datos: Azure SQL Database



- ① Creación y administración de bases SQL en la nube.
- ② Motor de SQL Server.
- ③ Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- ④ Escalabilidad y Seguridad.

Procesamiento de Datos: Azure SQL Database



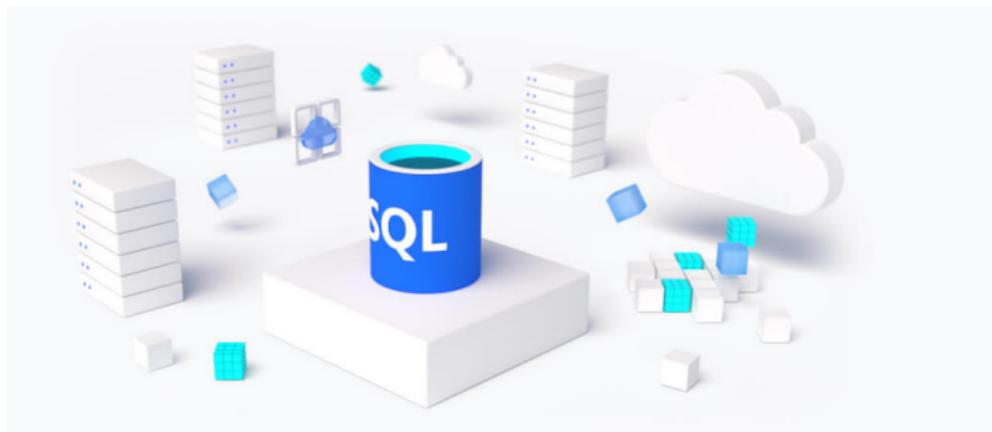
- ① Creación y administración de bases SQL en la nube.
- ② Motor de SQL Server.
- ③ Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- ④ Escalabilidad y Seguridad.

Procesamiento de Datos: Azure SQL Database



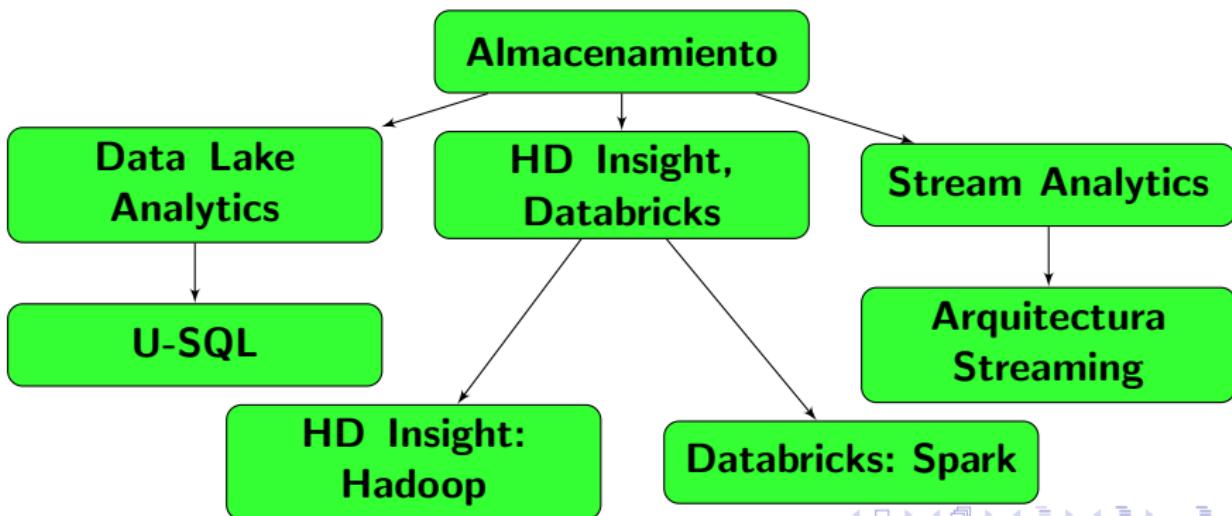
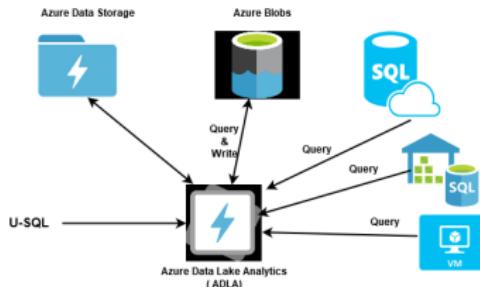
- ① Creación y administración de bases SQL en la nube.
- ② Motor de SQL Server.
- ③ Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- ④ Escalabilidad y Seguridad.

Procesamiento de Datos: Azure SQL Database

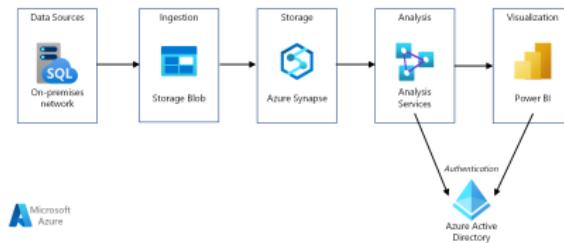


- I Creación y administración de bases SQL en la nube.
- II Motor de SQL Server.
- III Integrabilidad:
 - Azure Data Factory, Azure App Services, Azure Functions.
- IV Escalabilidad y Seguridad.

Almacenamiento en Paralelo



Data Lake y Data Warehouse



Data Lake y Warehouse

Data Lake

Datos (semi/no)
estructurados

Almacenamiento
y procesamiento

Warehouse

Datos estructurados

Análisis de negocios,
toma de decisiones

Características Principales



① Procesamiento de datos en nube.

- Creación, programación y orquestación de flujos de datos.
- Movimiento, transformación y carga.

② Integración de Azure Data Factory:

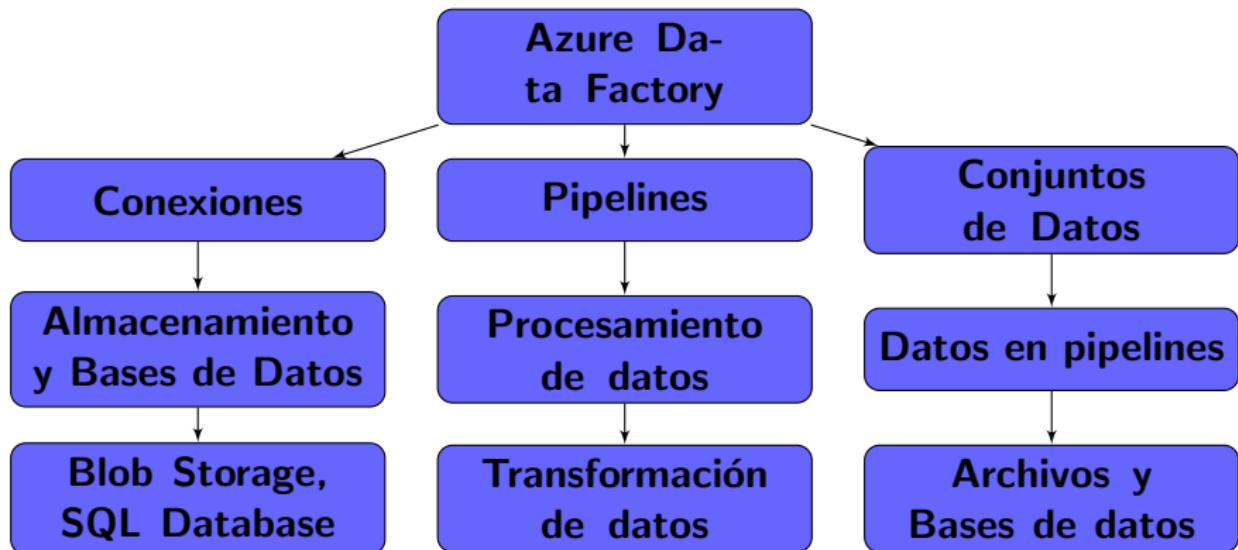
- Servicios de Azure (Databricks, Machine Learning, Stream Analytics).
- Visualización de datos (Power BI, Excel).

Características Principales

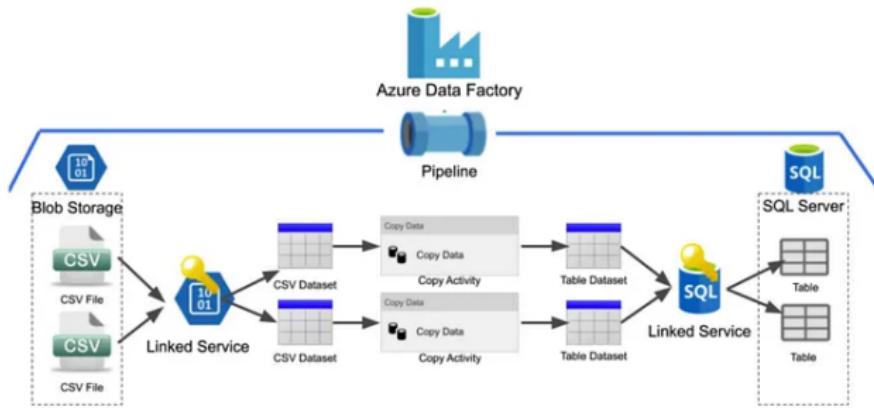


- ① Procesamiento de datos en nube.
 - Creación, programación y orquestación de flujos de datos.
 - Movimiento, transformación y carga.
- ② Integración de Azure Data Factory:
 - Servicios de Azure (Databricks, Machine Learning, Stream Analytics).
 - Visualización de datos (Power BI, Excel).

Usos

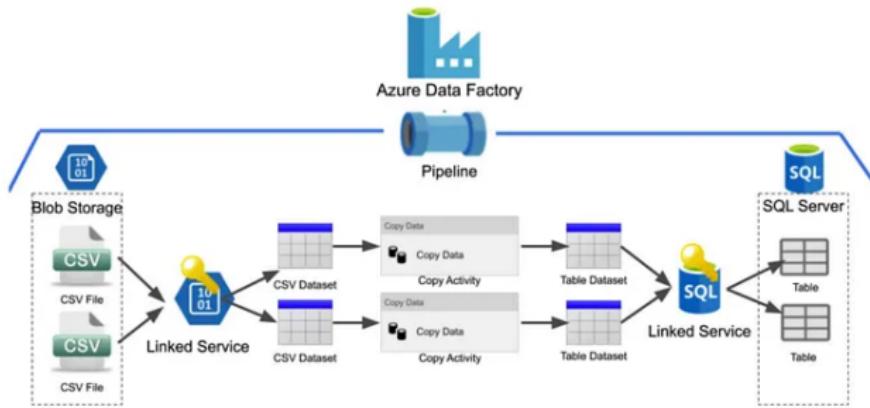


Data Factory como ETL



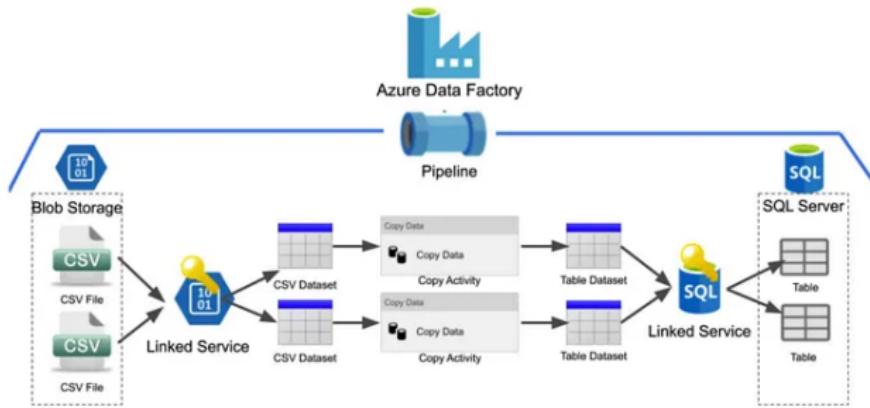
- ① Extracción: datos locales/en nube.
- ② Transformación/Limpieza: SQL, Python, herramientas visuales.
- ③ Carga: Bases, Datalakes.
 - *Linked Service:* conexión a un origen/destino de datos.

Data Factory como ETL



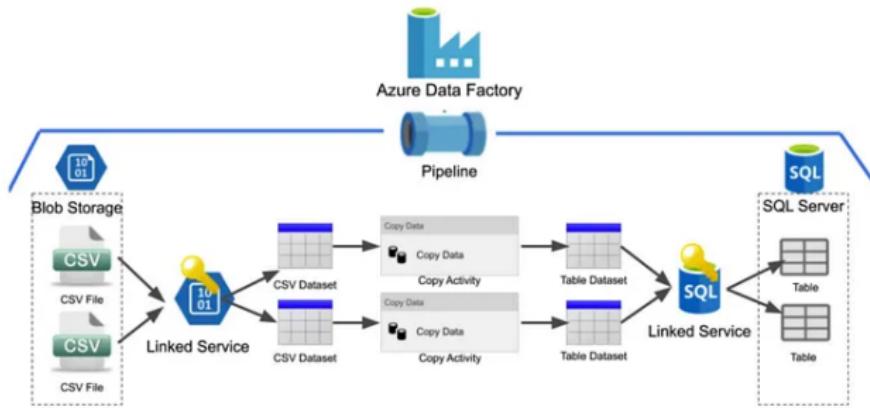
- ① Extracción: datos locales/en nube.
- ② Transformación/Limpieza: SQL, Python, herramientas visuales.
- ③ Carga: Bases, Datalakes.
 - *Linked Service:* conexión a un origen/destino de datos.

Data Factory como ETL



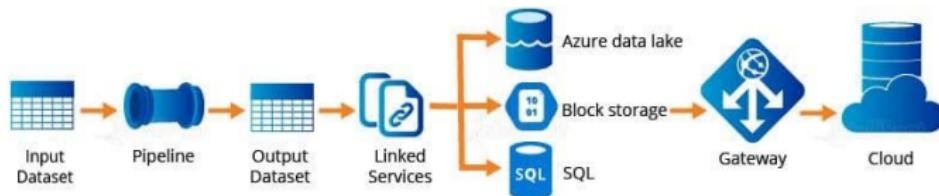
- ① Extracción: datos locales/en nube.
- ② Transformación/Limpieza: SQL, Python, herramientas visuales.
- ③ Carga: Bases, Datalakes.
 - *Linked Service*: conexión a un origen/destino de datos.

Data Factory como ETL



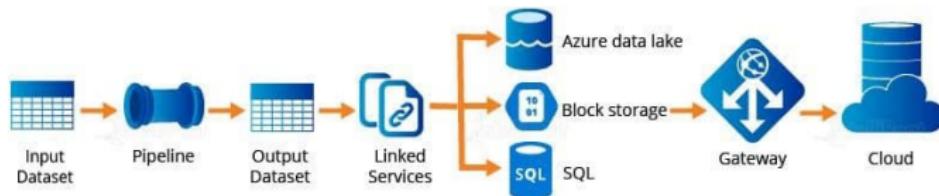
- ① Extracción: datos locales/en nube.
- ② Transformación/Limpieza: SQL, Python, herramientas visuales.
- ③ Carga: Bases, Datalakes.
 - *Linked Service*: conexión a un origen/destino de datos.

Datasets/Conjuntos de Datos



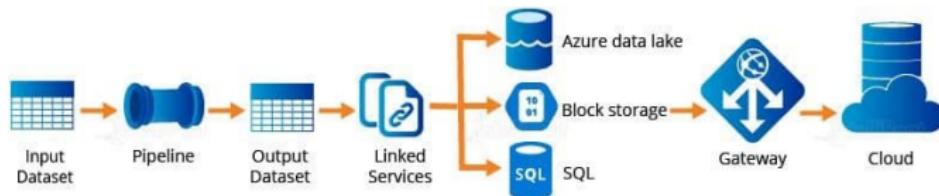
- ① Acceso a datos almacenados.
 - Uso en pipelines o dataflows.
- ② Fuentes/destinos: Blob Storage, Data Lake, SQL Server, etc.
- ③ Formatos:
 - Archivos de texto, excel, semiestructurados (JSON, XML), SQL (SQL server, MySQL).
 - Blob storage, Data Lake storage, Azure SQL Database.

Datasets/Conjuntos de Datos



- I Acceso a datos almacenados.
 - Uso en pipelines o dataflows.
- II Fuentes/destinos: Blob Storage, Data Lake, SQL Server, etc.
- III Formatos:
 - Archivos de texto, excel, semiestructurados (JSON, XML), SQL (SQL server, MySQL).
 - Blob storage, Data Lake storage, Azure SQL Database.

Datasets/Conjuntos de Datos



- I Acceso a datos almacenados.
 - Uso en pipelines o dataflows.
- II Fuentes/destinos: Blob Storage, Data Lake, SQL Server, etc.
- III Formatos:
 - Archivos de texto, excel, semiestructurados (JSON, XML), SQL (SQL server, MySQL).
 - Blob storage, Data Lake storage, Azure SQL Database.

Pipelines y Dataflows



① Pipelines:

- Orquestación de procesos ETL.
- Operaciones: dataflows, mover y transformar, anexar variables, borrar, etc.

② Dataflows:

- Procesamiento y transformación de grandes volúmenes de datos.
- Transformación datos entrada (interfaz gráfica).

Pipelines y Dataflows



① Pipelines:

- Orquestación de procesos ETL.
- Operaciones: dataflows, mover y transformar, anexar variables, borrar, etc.

② Dataflows:

- Procesamiento y transformación de grandes volúmenes de datos.
- Transformación datos entrada (interfaz gráfica).
 - A través de Python y SQL.

Actividades en Pipelines

① Mover y Transformar:

- Copiar datos, Dataflow (basado en SQL y Spark).

② Actividades Generales:

- If, ForEach, Set Variable, Until, Switch.

③ Actividades de Data Flow:

- Source, Derived Column, Filter, Aggregates, Join, Sink.

Actividades en Pipelines

① Mover y Transformar:

- Copiar datos, Dataflow (basado en SQL y Spark).

② Actividades Generales:

- If, ForEach, Set Variable, Until, Switch.

③ Actividades de Data Flow:

- Source, Derived Column, Filter, Aggregates, Join, Sink.

Actividades en Pipelines

I) Mover y Transformar:

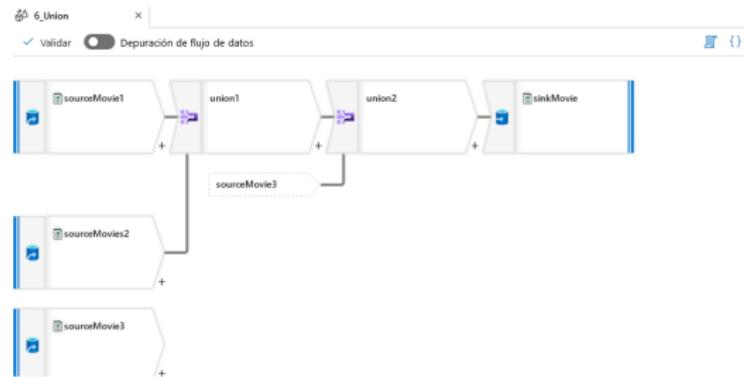
- Copiar datos, Dataflow (basado en SQL y Spark).

II) Actividades Generales:

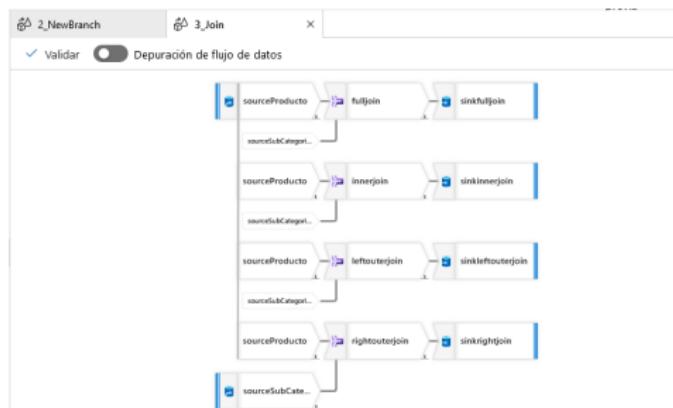
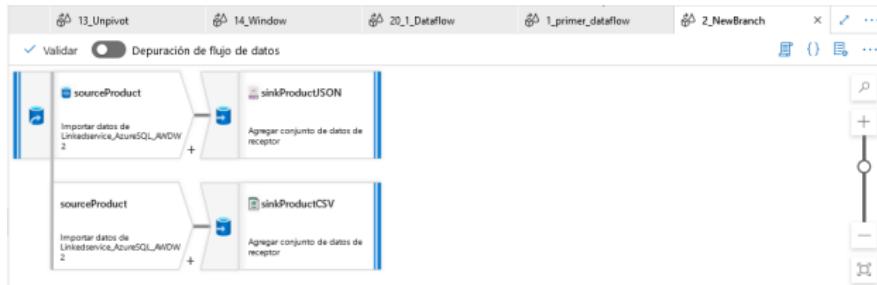
- If, ForEach, Set Variable, Until, Switch.

III) Actividades de Data Flow:

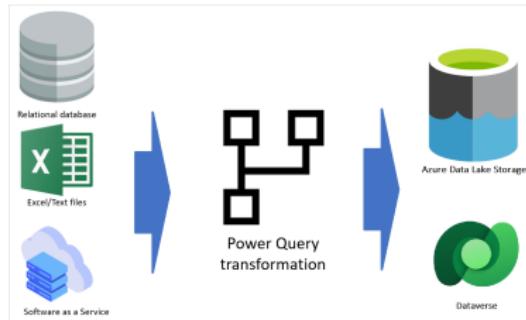
- Source, Derived Column, Filter, Aggregates, Join, Sink.



Dataflow en Azure Data Factory



Power Query



⑩ Actividad de transformación de datos.

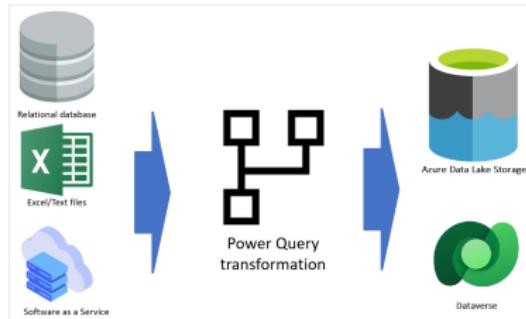
- Dentro de un Pipeline.
- Fuentes: archivos, DBs, APIs.

⑪ Uso:

- Definir actividad Power Query en Pipeline.

Definición de la actividad Power Query en el Pipeline

Power Query



① Actividad de transformación de datos.

- Dentro de un Pipeline.
- Fuentes: archivos, DBs, APIs.

② Uso:

- Definir actividad Power Query en Pipeline.
 - Ejemplos: Limpieza, transformación y agrupación de datos.